REMARKS

This Response is in reply to the Office Action mailed September 24, 2003, for which an unextended response is due December 24, 2003. Claims 1-21 and 24-26 are currently pending in the instant application.

In response to the Examiner's request, attached hereto are additional copies of all non-patent references previously submitted in the Information Disclosure Statement on July 14, 2003. For the Examiner's convenience, a copy of the PTO-1449 form submitted with the original references on July 14, 2003 is also included. Applicant respectfully requests the consideration of said references in the instant application.

Rejection of Claims 1-21 and 24-26 under 35 U.S.C. §112, ¶2

Claims 1-21 and 24-26 are rejected under 35 U.S.C. §112, second paragraph, as allegedly being indefinite due to the recitation of "adenovirus defective genome" in description of the first genic unit in claim 1. It is stated that the term "defective" is unclear, possibly referring to deletion of either the "minimal portion 600bp needed for replication and encapsidation" or the "encapsidation signal and a non-structural region." Applicant respectively traverses.

Applicant asserts that the description of the first genic unit as a "defective" adenovirus genome in claim 1 is unambiguous. Claim 1 recites that the first genic unit comprises an "adenovirus defective genome having the inverted terminal repeats in head-to-tail configuration, the encapsidation signal inactivated, and at least one of the non-structural regions inactivated" (emphasis added). Throughout the specification, the term "defective" is used to refer to an adenovirus genome that has been modified such that it is incapable, by itself, of either replication of viral DNA and/or encapsidation. For example, on page 3, lines 20-24, first generation vectors are described as generating virus that are "defective for the capability of replication" as a result of deletion of the E1 region, part of the "early" or "E" region of the viral genome that encodes polypeptides expressed in the first stages of infection. As such, the non-structural E1 proteins must be provided in *trans* to allow for proper replication of the modified adenovirus genome. Alternatively, "totally defective adenoviral vectors," described on page 4, lines 29-36, represent an adenovirus genome containing only a minimal portion of viral DNA, approximately 600 base pairs, required for replication and encapsidation upon supply of essential genes from a complementing cell line.

The first genic unit in claim 1 represents a modified adenovirus genome that is defined as "having" specific components: inverted terminal repeats, an inactivated encapsidation signal and

at least one inactivated non-structural region. The encapsidation signal is always non-functional in the adenovirus genome of the first genic unit, making said genome "defective" due to its inability to assemble viral particles by itself. Additionally, "at least one" of the non-structural genes encompassing the viral DNA E region is inactivated. Applicant asserts that the definiteness of the "defective" adenovirus genome of the first genic unit comes from the specific recitation of its components, including both the necessary components (*e.g.*, inverted terminal repeats) and modified non-functional components (*e.g.*, the inactivated encapsidation signal and at least one inactivated non-structural region), which in combination render the genome "defective."

Based on the arguments presented above, Applicant asserts that claim 1 and its depending claims meet the standard set out in 35 U.S.C. §112, second paragraph, by particularly pointing out and distinctly claiming the subject matter which Applicant regards as the invention, particularly to one of skill in the pertinent art. Accordingly, Applicant respectfully requests reconsideration and withdrawal of the instant rejection.

Rejection of Claims 8, 9, 18-21 under 35 U.S.C. §112, ¶2

Claims 8, 9 and 18-21 are rejected under 35 U.S.C. §112, second paragraph, as allegedly being indefinite due to the recitation of the phrase "totally or partially constituted" in description of the first or second genic unit. It is stated that the cited phrase is unclear because it does not definitively describe how much of the first or second genic unit is comprised of human adenovirus sequences and how much may be derived from other viral genic units. Applicant respectfully traverses. Applicant asserts that claims 18-21 are not indefinite in recitation of the phrase "totally or partially constituted" because one of ordinary skill in the art could readily appreciate how much of either the first or second genic unit could be comprised of sequences from a single adenovirus species (*e.g.*, human adenovirus sequences).

The instant specification teaches a method of producing helper dependent adenoviral vectors utilizing cells containing a first and second genic unit which, together, express the appropriate adenoviral proteins needed to complement the production of said vector in the absence of helper virus. Indeed, claim 1 specifically states that it is the presence of the first and second genic units, together, which "enable[s] the production" of the helper dependent adenoviral vector. As explained in the specification, the genic information encoding the essential adenoviral proteins provided by the claimed cells is divided between two genic units in order to more strictly regulate the expression of said products. Importantly, however, the precise origin of the sequences comprising the first and second genic units is less critical than the function of their

encoded products in the novel process disclosed. The genic units are designed to satisfactorily express a function capable of complementing the inactivated/absent function(s) of the helper dependent adenovirus vector. If a different function is meant to be served, for example, by according a different phenotype (*e.g.*, supply of a heterologous capsid protein) or by effecting the production of a different species (*e.g.*, swine), it is within the realm of one skilled in the art to design specific hybrid constructs that will function in the disclosed process in pursuit of that goal. In such a case, the composition of the two genic units would be dictated by the desired phenotype, serotype or function of the ultimate helper dependent adenovirus vector. Thus, the present invention does not hinge on any specific characteristic or feature of any one adenovirus species (*e.g.*, an adenovirus that infects humans, as exemplified). Applicant maintains that the production of each and every possible adenovirus, including various species (native or chimeric forms), is not necessary to satisfy the requirements of 35 U.S.C. §112, second paragraph. What is disclosed is a generalized process of how to construct such adenovirus in a more regulated manner through the supply of essential proteins through these genic units. Based on these arguments, Applicant asserts that claims 18 and 20, and the dependent claims 19 and 21, particularly point out and distinctly claim the subject matter regarded as the invention, especially to one of skill in the pertinent art. Applicant, therefore, respectfully requests reconsideration and withdrawal of the instant 35 U.S.C. §112, second paragraph, rejection of said claims.

In regards to claim 8 and claim 9, since neither claim contains the phrase "totally or partially constituted," and the explanation of the instant indefiniteness rejection does not appear to apply to said claims, Applicant respectfully requests withdrawal of the instant rejection to claims 8 and 9.

Rejection of Claims 1-21 and 24-26 under 35 U.S.C. §112, ¶1

Claims 1-21 and 24-26 are rejected under 35 U.S.C. §112, first paragraph, as allegedly failing to enable one of ordinary skill in the art to practice the invention commensurate in scope with the cited claims. The instant enablement rejection is based on the fact that the current claims are not limited to human adenoviruses, and thus, can include all adenovirus from other species. Since adenovirus from alternate species can differ at the nucleotide level, it is alleged that one of skill in the art would lack the ability to both "manipulate a comparable sequence at the molecular level" and "predict with reasonable success the structure and function of expressed proteins." Applicant respectfully traverses.

As stated above, Applicant has disclosed a process for the production of helper dependent adenoviral vectors independent of a helper virus. The role of said helper virus is replaced by the

function of two genic units contained within the claimed helper cells. It is within the expertise of one skilled in the art to identify those adenoviral components which may need to be supplied by the helper cells, via the introduced genic units, in order to provide complementing activity. For example, without resorting to undue experimentation, said skilled individual could readily identify those components required for replication and/or particle assembly by deleting regions in a chosen wild-type adenovirus genome and determining whether proper replication and/or virion assembly occurs without providing said deleted regions in *trans*. Similarly, using routine molecular biology techniques, one of skill in the art could manipulate essential regions of adenovirus genomes of alternate species/serotypes for inclusion within the first/second genic units described in claim 1.

Further supporting Applicant's argument, at the time the instant application was originally filed, November 6, 1998, many complete sequences for adenovirus genomes of alternate species were known (see Table I of Davison, A.J. et al., 2003, *J. Gen. Virol.* 84:2895-2908; a copy of which is attached hereto as Exhibit A), facilitating their manipulation and ensuring a likelihood of successful expression in the genic units described herein. One of skill in the art is certainly cognizant of the fact that while there can be some sequence variation among adenovirus that infect various species, there is a reasonable consistency among the generalized functions of certain adenoviral gene components; and irrespective of any consistency, the distinct function of different adenoviral components are well appreciated in the art, particularly those executing the primary functions of DNA replication, DNA encapsidation and virion formation. By comparing the known sequences of different adenovirus genomes available at the time the instant application was filed, one of skill in the art could have easily identified those sequences essential for helper dependent adenoviral vector production via the process taught by Applicant.

Based on these arguments, Applicant asserts that claims 1-21 and 24-26 enable one of skill in the pertinent art to practice the invention commensurate in scope with said claims without undue experimentation. Thus, Applicant respectfully requests reconsideration and withdrawal of the instant 35 U.S.C. §112, first paragraph, rejection of said claims.

In view of the comments herein, Applicant respectfully takes the position that claims 1-21 and 24-26 are in proper form for allowance. The Examiner is invited to contact the undersigned attorney if clarification is required on any aspect of this response, or if any of the claims are considered to require further amendment to be placed in condition for allowance after entry of this response.

Respectfully submitted,

By Laura M. Ginkel

Laura M. Ginkel
Reg. No. 51,737
Attorney for Applicant

MERCK & CO., INC.
P.O. Box 2000
Rahway, New Jersey 07065-0907
(732) 594-1932

Date: Dec. 19, 2003

Journal of General Virology (2003), 84, 2895–2908 DOI 10.1099/vir.0.19497-0

Review

Genetic content and evolution of adenoviruses

Andrew J. Davison,[1] Mária Benkő[2] and Balázs Harrach[2]

Correspondence
Andrew Davison
a.davison@vir.gla.ac.uk

[1]MRC Virology Unit, Institute of Virology, Church Street, Glasgow G11 5JR, UK

[2]Veterinary Medical Research Institute, Hungarian Academy of Sciences, H-1581 Budapest, Hungary

This review provides an update of the genetic content, phylogeny and evolution of the family *Adenoviridae*. An appraisal of the condition of adenovirus genomics highlights the need to ensur that public sequence information is interpreted accurately. To this end, all complete genome sequences available have been reannotated. Adenoviruses fall into four recognized genera, plus possibly a fifth, which have apparently evolved with their vertebrate hosts, but have also engaged in a number of interspecies transmission events. Genes inherited by all modern adenoviruses from their common ancestor are located centrally in the genome and are involved in replication and packaging of viral DNA and formation and structure of the virion. Additional niche-specific genes have accumulated in each lineage, mostly near the genome termini. Capture and duplication of genes in the setting of a 'leader–exon structure', which results from widespread use of splicing, appear to have been central to adenovirus evolution. The antiquity of the pre-vertebrate lineages that ultimately gave rise to the *Adenoviridae* is illustrated by morphological similarities between adenoviruses and bacteriophages, and by use of a protein-primed DNA replication strategy by adenoviruses, certain bacteria and bacteriophages, and linear plasmids of fungi and plants.

INTRODUCTION

The purpose of this review is to provide a comprehensive update of the genetic content, phylogeny and evolution of the family *Adenoviridae*, whose members infect hosts throughout the vertebrates (Russell & Benkő, 1999). This area has frequently taken a back seat to studies of the interactions of selected human adenovirus proteins with cellular processes and, more recently, to the use of adenoviruses as vectors. Our aim is to redress this imbalance by bringing together published information and by offering fresh insights into the genomics of the family as a whole. It is not our purpose to deal in broad scope with other areas, such as the expression, functions and interactions of adenovirus gene products and the ways in which host defences are manipulated during the infectious cycle. These have been reviewed recently by Russell (2000) and Shenk (2001) for human adenoviruses.

Phylogeny, classification and genetic rganization

Members of the family *Adenoviridae* are non-enveloped, icosahedral viruses that replicate in the nucleus. Their linear, double-stranded DNA molecules are 26–45 kbp in size and rank as medium-sized among the DNA viruses. The genomes are characterized by an inverted terminal repeat (ITR) ranging in size from 36 to over 200 bp, and the 5′ ends are linked to a terminal protein (TP). Phylogenetic relationships among a large number of adenoviruses infecting vertebrates from fish to humans are shown in Fig. 1. The major clades (groups of viruses sharing a common ancestor) correspond to the four accepted genera plus a fifth genus that is likely to be added in due course (Benkő *et al.*, 2002; Harrach & Benkő, 1998). Two genera (*Mastadenovirus* and *Aviadenovirus*) originate from mammals or birds, respectively, and the other two genera (*Atadenovirus* and *Siadenovirus*) have a broader range of hosts. Atadenoviruses were named as a consequence of the bias of their genomes towards high A+T content (Benkő & Harrach, 1998; Both, 2002a), and infect various ruminant, avian and reptilian hosts, as well as a marsupial. The two known siadenoviruses were isolated from birds and a frog (Davison & Harrach, 2002). The only confirmed fish adenovirus falls into the fifth clade. Within each genus, viruses are grouped into species, which are named from the host and supplemented with letters of the alphabet, as listed in Table 1 (Benkő *et al.*, 2000). Host origin is only one of several criteria used to demarcate the species taxon. Thus, for example, known chimpanzee adenoviruses are classified into human adenovirus species. In the case of human adenoviruses, the present species correspond to groups or subgenera defined previously (Bailey & Mautner, 1994; Wadell, 1984). Questions of classification into species are not yet resolved for many non-human adenoviruses.

The GenBank accession numbers of the Third Party Annotations of adenovirus genome sequences reported in this paper are listed in Table 1.

Published ahead of print on 13 August 2003 as DOI 10.1099/vir.0.19497-0

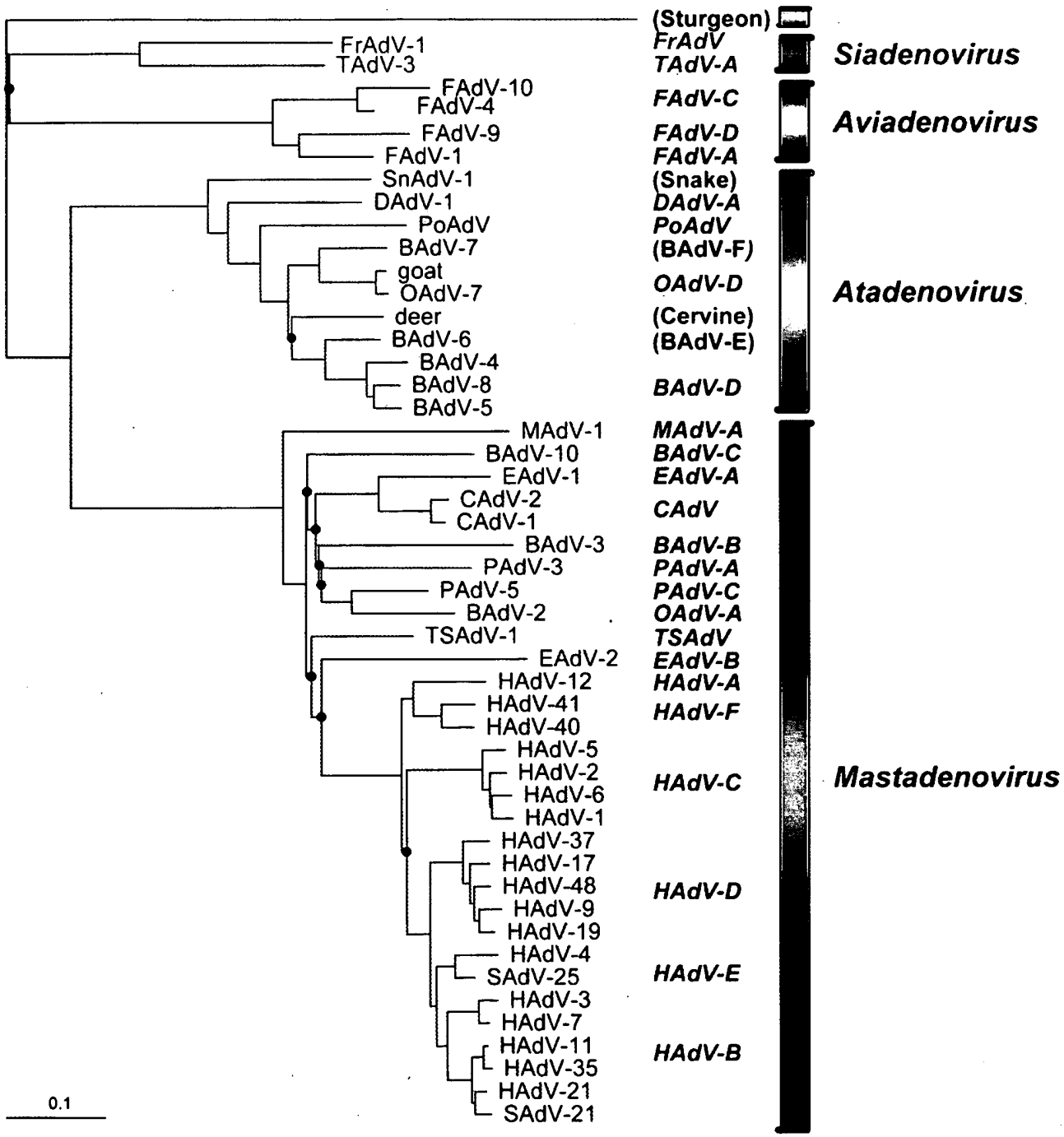


Fig. 1. Distance tree summarizing the phylogeny of adenovirus hexon genes. Members of the various genera are indicated in different colours, and viruses that belong to the same species are grouped by light-green ovals. Abbreviations of virus names are indicated at the ends of the branches, with species names listed to the right (recognized species in italics): B, bovine; C, canine; D, duck; E, equine; F, fowl; Fr, frog; H, human; M, murine; O, ovine; P, porcine; Po, possum; Sn, snake; T, turkey; and TS, tree shrew. The distance matrix was calculated from amino acid sequences for hexon available in GenBank (in some cases combined from different partial entries) and from our unpublished sequences. The PROTDIST (Dayhoff PAM 001 matrix) and FITCH (global rearrangements) programs of the PHYLIP package, version 3.6, were used. The tree was rooted by specifying sturgeon adenovirus as the outgroup and displayed using TREEVIEW (Page, 1996). Bootstrap values of less than 80/100 (not including branches within species) are indicated by small, filled circles.

Tabl 1. Available s quences for complete ad novirus genomes

GENUS Speci s Serotype	Abbreviation	Accession	Siz (bp)	R f rence	TPA
MASTADENOVIRUS					
Human adenovirus A	***HAdV-A***				
Human adenovirus 12	HAdV-12	X73487	34125	Sprengel *et al.* (1994)	BK000405
Human adenovirus B	***HAdV-B***				
Human adenovirus 11	HAdV-11	AY163756*	34794	Stone *et al.* (2003)	BK001453
Simian adenovirus 21	SAdV-21	AR101858	35524	Patent application (2000)	BK000412
Human adenovirus C	***HAdV-C***				
Human adenovirus 2	HAdV-2	J01917	35937	Roberts *et al.* (1986)	BK000407
Human adenovirus 5	HAdV-5	M73260	35935	Chroboczek *et al.* (1992)	BK000408
Human adenovirus D	***HAdV-D***				
Human adenovirus 17	HAdV-17	AF108105	35100	Chillon *et al.* (1999)	BK000406
Human adenovirus E	***HAdV-E***				
Simian adenovirus 25	SAdV-25	AR101859†	36519	Patent application (2000)	BK000413
Human adenovirus F	***HAdV-F***				
Human adenovirus 40	HAdV-40	L19443	34214	Davison *et al.* (1993)	¶
Ovine adenovirus A	***OAdV-B***				
Bovine adenovirus 2	BAdV-2	AF252854	33034	Unpublished (2000)	BK000400
Porcine adenovirus C	***PAdV-C***				
Porcine adenovirus 5	PAdV-5	AF289262	32621	Nagy *et al.* (2001)	BK000411
B vine adenovirus B	***BAdV-B***				
Bovine adenovirus 3	BAdV-3	AF030154	34446	Reddy *et al.* (1998b)	BK000401
Canine adenovirus	***CAdV***				
Canine adenovirus 1	CAdV-1	Y07760‡	30536	Morrison *et al.* (1997)	BK000402
Canine adenovirus 2	CAdV-2	U77082	31323	Unpublished (1996)	BK000403
Porcine adenovirus A	***PAdV-A***				
Porcine adenovirus 3	PAdV-3	AF083132§	34094	Reddy *et al.* (1998a)	BK000410
Murine adenovirus A	***MAdV-A***				
Murine adenovirus 1	MAdV-1	NC_000942	30944	Meissner *et al.* (1997)	BK000415
Tree shrew adenovirus	***TSAdV***				
Tree shrew adenovirus 1	TSAdV-1	AF258784	33501	Schöndorf *et al.* (2003)	BK001455
ATADENOVIRUS					
Duck adenovirus A	***DAdV-A***				
Duck adenovirus 1	DAdV-1	Y09598	33213	Hess *et al.* (1997)	BK000404
Ovine adenovirus D	***OAdV-D***				
Ovine adenovirus 7	OAdV-7	U40839	29574	Vrati *et al.* (1996)	¶
B vine adenovirus D	***BAdV-D***				
Bovine adenovirus 4	BAdV-4	AF036092	31300	Unpublished (1997)	¶
AVIADENOVIRUS					
Fowl adenovirus A	***FAdV-A***				
Fowl adenovirus 1	FAdV-1	U46933	43804	Chiocca *et al.* (1996)	BK001452
F wl adenovirus D	***FAdV-D***				
Fowl adenovirus 9	FAdV-9	AF083975	45063	Ojkic & Nagy (2000)	BK001451
SIADENOVIRUS					
Turkey adenovirus A	***TAdV-A***				
Turkey adenovirus 3	TAdV-3	AF074946	26263	Pitcovski *et al.* (1998)	BK001454
Frog adenovirus	***FrAdV***				
Frog adenovirus 1	FrAdV-1	AF224336	26163	Davison *et al.* (2000)	¶

*Another sequence is available for this strain (AF532578; Mei *et al.*, 2003).

†Another sequence was derived subsequently for this serotype by the same group (AF394196; Farina *et al.*, 2001). It is not clear whether this is from the same strain, but in any case it is substantially inferior in quality, containing 59 ambiguous nucleotides and several frameshifts.

‡Another sequence is available for a strain of this serotype (U55001; unpublished, 1996). This differs from the sequence given in the Table at 53 locations, as well as in a deletion of part of the E3 region. None of these differences indicates frameshift errors.

§A duplicated accession for a passaged derivative of this strain is available (AB026117 and AJ237815; unpublished, 1999). This sequence differs from that given in the Table at 33 locations. Eight of these differences indicate frameshift errors in one or other sequence.

¶TPAs were not submitted since the authors' original accessions have been updated directly.

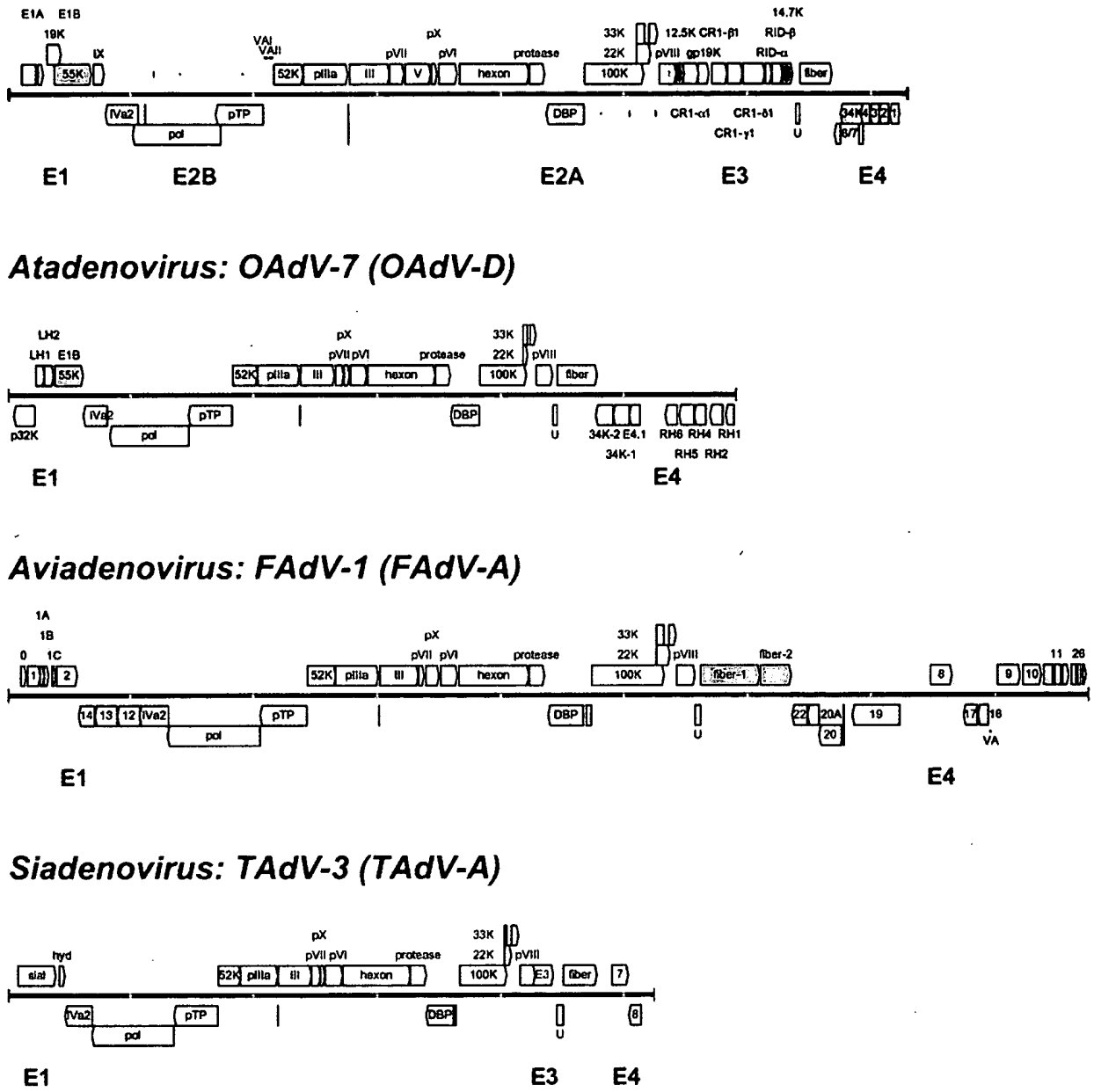


Fig. 2. Gene layout in representatives of the four adenovirus genera recognized. Each genome is represented by a central blue horizontal line marked at 5 kbp intervals. Protein-encoding regions are shown as arrows. Cyan denotes genus-common genes, other colours indicate related genus-specific genes (E1B 55K and E4 34K are related between the *Mastadenovirus* and *Atadenovirus* genera), and genus-specific genes that lack relatives are white. E1 and E4 are marked in all genera and E3 in the *Mastadenovirus* and *Siadenovirus* genera, and these regions are shaded violet. E2A and E2B genes are present in all genera, and these regions are indicated for the *Mastadenovirus* genus. Untranslated leader exons in the *Mastadenovirus* genus are shown for late, E2, E3 and E4 genes, with early transcripts in red, late transcripts in blue and 5′-terminal exons broader than internal exons. The VA RNA genes in the *Mastadenovirus* and *Aviadenovirus* genera are also included.

Gene arrangements for representative members of the adenovirus genera recognized are illustrated in Fig. 2. The upper panel shows the layout for a human adenovirus species (specifically, a chimpanzee virus). In the discussion below, we use the same term (e.g. pX) to apply to gene or protein, as indicated by the context, with proteins that are produced as precursors for cleavage by the viral protease prefaced by a 'p'. In outline, genomes of human adenovirus species consist of a central block of rightward-oriented late genes from 52K to fiber, interrupted on the same strand by a

block of early genes in the E3 region and on the opposing strand by E2 genes in the form of DNA-binding protein (DBP) in the E2A region and pre-terminal protein (pTP) and DNA polymerase (pol) in the E2B region. The right-terminal region is occupied by E4 genes and the left-terminal region by early E1A and E1B genes plus two intermediate genes (IX and IVa2).

Splicing

Splicing was first discovered in human adenoviruses as the mechanism by which several late genes are expressed from a shared non-translated tripartite leader (Berget *et al.*, 1977; Chow *et al.*, 1977). It has since become clear that splicing is involved in expression of most human adenovirus genes (Akusjärvi *et al.*, 1986). Thus, all members of the array of late genes (which are largely involved in virion formation and structure), from 52K rightwards to pVIII, and also fiber, are spliced from the tripartite leader (Fig. 2). Similarly, early genes and one intermediate gene are spliced. These include: E1A; DBP, which is spliced from a leader consisting of two untranslated exons controlled by the E2 promoter (expression also occurs from a separate late promoter; Fig. 2); pTP and pol, which are spliced from a leader consisting of two untranslated and one translated exons controlled by the E2 promoter (Fig. 2); IVa2, an intermediate gene that is spliced from a short coding leader within pol; and the E3 and E4 groups of genes, each of which is initiated from a single promoter so that the proximal gene is not spliced whereas others are expressed by splicing from the relevant untranslated leader (Fig. 2). Transcription of E1A, E1B, E3 and E4, and to a lesser extent other genes, has been perceived historically as more complex, potentially giving rise via alternative splicing to a large catalogue of RNAs and proteins, some of which have detectable functions in *in vitro* assays (Akusjärvi *et al.*, 1986; Chow *et al.*, 1979; Tigges & Raskas, 1984; Virtanen *et al.*, 1984; Wold & Gooding, 1991). However, transcript mapping studies have largely been carried out on a single human adenovirus species (HAdV-C, represented by HAdV-2 and HAdV-5), and many of the alternative splice sites are not conserved even among the human adenovirus species. Indeed, an overview of mastadenovirus genomes now forces us to question whether these evolutionarily ephemeral genes have meaningful roles during the natural life cycle. On this basis, we have omitted from Fig. 2 any coding regions specified by alternative splicing.

Splicing is not universal in adenoviruses: in the human viruses, E1B 19K, E1B 55K and IX are single exon genes. Nonetheless, a pattern in which a common leader is spliced to any one of several downstream protein-encoding exons is evident in the late, E2, E3 and E4 regions. This simple layout, termed here the 'leader–exon structure', facilitates the evolution of new genes by capture or duplication, merely requiring the presence of a splice acceptor site upstream from a newly inserted protein-encoding region. This strategy has apparently functioned throughout adenovirus evolution and is considered in further detail below. As far as can be deduced from sequence comparisons, the general splicing patterns of the late and E2 genes is common to all adenoviruses. Thus, late genes are spliced from an untranslated leader, whether tripartite, as in mastadenoviruses and atadenoviruses (Khatri & Both, 1998), or bipartite, as in aviadenoviruses (Sheppard *et al.*, 1998). Also, in all adenoviruses, the initiation codon for pTP is located in a short exon between pIIIa and III (Fig. 2). This initiation codon is also utilized for expression of pol in the mastadenoviruses, but apparently not in the other genera, where the short exon is probably still spliced to the main coding exon, but pol translation is initiated in the main exon. These differences reflect minor evolutionary tinkering; in this case, the change of register in splice sites to allow the leader in the leader–exon structure to switch between being non-coding and coding. Whichever situation pertains, the structure lends itself well to evolution of new genes. Differences between the genera are also evident in splicing of DBP. This gene is probably spliced from upstream non-coding exons in all adenoviruses, but in aviadenoviruses and siadenoviruses it has accrued an additional coding exon upstream from the main exon.

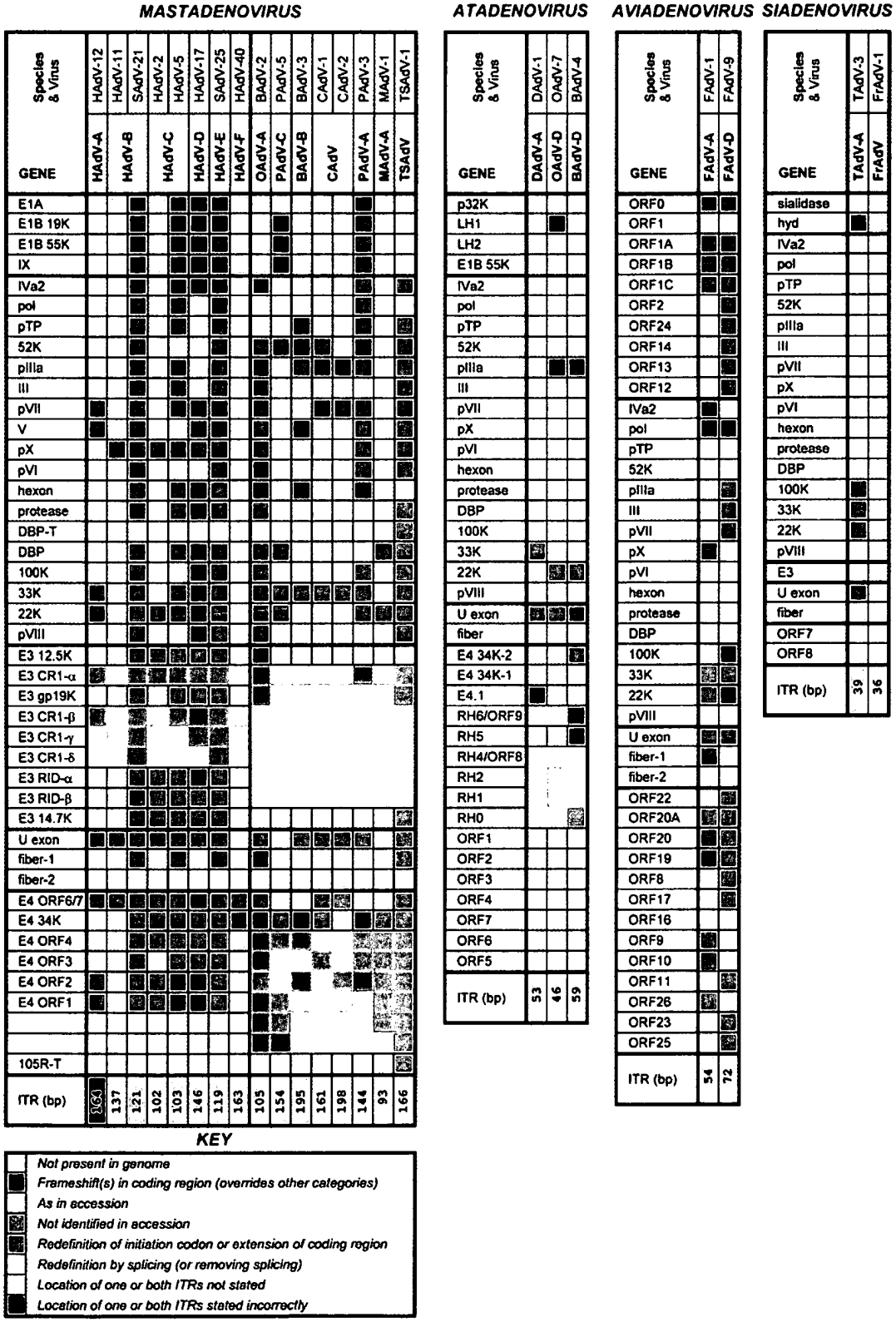
Genome sequences

We conducted a thorough analysis of complete adenovirus genome sequences deposited in GenBank (http://www.ncbi.nlm.nih.gov/), utilizing current knowledge of adenovirus gene content. The sequences are derived from 23 serotypes encompassed by the four genera recognized, and are listed in Table 1. Our conclusions on the status of the entries are summarized in Table 2. Shortcomings in interpretation and annotation are obvious in several respects, and often belie a much higher quality of analysis in the primary publications. Firstly, annotations may be incomplete and in some cases totally absent. Secondly, the coordinates of some protein-encoding regions may be incorrect owing to neglect of comparative data or splicing. Thirdly, a sizeable subset of genes may exhibit frameshifts, most, if not all, due probably to sequencing errors. Indeed, several errors in the HAdV-5 sequence were demonstrated experimentally by Dix & Leppard (1992) and have been confirmed by us. Errors outside protein-encoding regions may also be apparent, for example, in differences between the supposedly identical ITRs of TSAdV-1 and also of DAdV-1. Fourthly, updated analysis may reveal previously unrecognized genes, some with unequivocal cellular counterparts. New examples from this category are incorporated into the discussion below.

These findings imply that access to reliable adenovirus genome data is compromised. As a result, the amino acid sequences of many adenovirus proteins are unavailable, and some of those that are available are incorrect. This situation may be mitigated by separate and more extensive annotation of fragments of certain genomes. However, the extent of this amelioration is partial, uneven and laborious to ascertain for individual viruses. For example, although GenBank contains entries for only eight HAdV-5 proteins derived from the

Tabl 2. Accuracy and annotation status of complet adenovirus genom accessions

Cells lacking black borders indicate regions where orthology is difficult to establish and is therefore not implied. This is because the genes therein are highly variable in number and sequence and in some cases belong to families of related genes. There is little or no similarity between certain primate and non-primate mastadenovirus genes in E3 and E4, and the nomenclature shown applies only to the former. The nomenclature of E3 CR1 genes, which are present only in primate mastadenoviruses, is based on order in the genome.



genome sequence, entries for the majority of proteins are accessible via annotations of genome fragments. In contrast, amino acid sequences are accessible only for the six genes that are annotated in the complete TSAdV-1 genome, as entries for fragments provide no additional information. This evaluation of the dismal condition of adenovirus genomics also applies to entries in the GenBank Reference Sequence library (RefSeq, http://www.ncbi.nlm.nih.gov/PMGifs/Genomes/viruses.html), which is intended as a means of avoiding fossilization of sequence interpretations. These entries are presently very similar to the originals, though some have been updated sparingly and with varying success by the gene prediction program GENEMARK (http://opal.biology.gatech.edu/GeneMark/) (Borodovsky & McIninch, 1993). Accession numbers for our annotations of adenovirus genomes are listed in Table 1, and are available under the GenBank Third Party Annotation scheme (TPA, http://www.ncbi.nlm.nih.gov/Genbank/tpa.html). Although questions remain concerning expression of certain regions, we consider that these entries are the best available at present. We conclude this section by noting that the lack of a readily updated data resource is hindering systematic advancement of comparative adenovirology.

The nucleotide composition properties of adenovirus genomes have attracted interest, with a focus on the atadenoviruses (Benkő & Harrach, 1998; Farkas *et al.*, 2002). Wide variation in nucleotide composition is characteristic of many groups of organisms, including adenoviruses, but the driving evolutionary forces remain elusive. Davison *et al.* (2000) highlighted the fact that certain adenovirus genomes (mastadenoviruses CAdV-1 and CAdV-2, atadenovirus OAdV-7 and siadenoviruses TAdV-3 and FrAdV-1) are depleted throughout in the CG dinucleotide. In cellular genomes, this evolutionary phenomenon is thought to result from methylation of the cytosine residue in the CG dinucleotide, followed by spontaneous deamination to TG and fixation by DNA replication. CG depletion is also a feature of certain herpesviruses and has been attributed to methylation of latent genomes resident in dividing cell populations (Honess *et al.*, 1989). This parallel yields provocative indicators to unexplored aspects of adenovirus biology.

Genus-common genes

The principal means of assigning genes across the *Adenoviridae* is comparative analysis, which involves identifying conserved protein-encoding regions. Broadly, we class genes with homologues in all genera as 'genus-common genes' and all others as 'genus-specific genes'.

There are 16 clearly defined genus-common genes. We presume that these were inherited from a common ancestral adenovirus, in which they were all expressed by splicing. Their primary functions are in DNA replication (pol, pTP and DBP), DNA encapsidation (52K and IVa2) and formation and structure of the virion (pIIIa, III, pVII, pX, pVI, hexon, protease, 100K, 33K, pVIII and fiber). Two additional protein-encoding regions may be added. One, 22K, originates from lack of splicing in 33K, and thus the N-terminal sequence of the protein is identical to that of 33K but the C-terminal sequence is encoded by the 33K intron. However, the putative 22K-encoding region of atadenoviruses does not extend through the intron, and thus its inclusion for this genus is tentative. The second addition is the U exon, which extends from an initiation codon to a splice donor site and is regulated by a minor late promoter (Chow *et al.*, 1979; Davison *et al.*, 1993, 2000). The downstream exons spliced to this exon have not yet been identified. They may correspond to IVa2, pol, pTP or DBP and yield differentially spliced forms of one or more of these genes encoding N-terminally extended proteins. The U exon appears to be a genus-common feature that has been lost in certain mastadenoviruses (PAdV-5 and MAdV-1).

Candidate splice sites for the majority of genus-common genes may be predicted from DNA sequences. However, particular uncertainty should be expressed about two. One gene, 33K, is poorly conserved among genera and identification of the splice sites is tentative in some cases. Also, the first exon is not visible in DAdV-1. The second problem concerns IVa2, which is involved in transcriptional activation of the late promoter and in capsid assembly and DNA packaging (Zhang & Imperiale, 2003). In mastadenoviruses, conservation of the region upstream from the first ATG codon supports splicing from an upstream protein-encoding exon. This is in accord with data that demonstrate splicing of the main coding region from a short leader close upstream, such that the five N-terminal residues of IVa2 correspond to residues within pol (Baker & Ziff, 1981; Reddy *et al.*, 1998a). However, the splice sites are not conserved throughout the mastadenoviruses. In atadenoviruses, mapping data for OAdV-7 indicate that IVa2 is not spliced (Khatri & Both, 1998), but this is at variance with conservation of the region upstream from the initiation codon, and again points to the existence of an upstream coding exon. We conclude that the transcriptional pattern of IVa2, which is likely to have been inherited by all genera, is worth further investigation.

Several adenovirus proteins (genus-common pTP, pIIIa, pVII, pX, pVI and pVIII, and genus-specific p32K of atadenoviruses) are cleaved by the protease in steps that are essential for the synthesis of infectious virus (Weber, 1995). Reported consensus cleavage sites may be summarized as (M/L/I)XGG'X and (M/L/I/N/Q)X(A/G)X'G (Anderson, 1990; Farkas *et al.*, 2002; Ruzindana-Umunyana *et al.*, 2002; Vrati *et al.*, 1996; Webster *et al.*, 1989). Since the protease is a genus-common gene, proteolytic maturation is evidently an ancient feature of the adenovirus replicative cycle. As an example, Fig. 3 shows putative cleavage sites for all of the sequenced viruses in pX, which, along with pVII and V, generates the core proteins of the virion. The primary translation product consists of an extended, basic N-terminal region linked to a short, basic C-terminal region via a highly conserved hydrophobic domain. This protein is cleaved by the protease between the basic N-terminal and

Fig. 3. Alignment of the complete predicted primary amino acid sequences of pX from the four adenovirus genera recognized. The overall alignment is centred on the conserved putative transmembrane domain, shown in blue. Predicted or confirmed protease cleavage sites are indicated by red vertical lines. Sites conform to (M/L/I/V/F)XGG'X and (M/L/I/V/N/Q)X(A/G)X'G, and are slightly less specific at the first residue than those described in the literature [(M/L/I)XGG'X and (M/L/I/N/Q)X(A/G)X'G].

hydrophobic domains (and in some viruses at a second site nearer the N terminus) to give rise to a core protein (X or μ) that is closely associated with the virion DNA. An incidental, but provocative, observation is that unprocessed pX in all adenoviruses has the primary sequence features of a class II membrane protein.

Genus-specific genes

Most genus-specific genes are located near the ends of the genome. These regions are termed E1 and E4 for human adenoviruses, and, despite the general lack of genetic similarity between genera in these regions, we utilize this nomenclature for all adenoviruses. In mastadenoviruses and siadenoviruses, genus-specific genes are also located in the E3 region, between pVIII and the U exon, and, in addition, the mastadenoviruses contain a single genus-specific gene (V) between pVII and pX. This general location pattern for genus-specific genes is not restricted to adenoviruses. It is also a feature of other linear, double-stranded DNA genomes, such as the *Herpesviridae* (McGeoch & Davison, 1999) and *Poxviridae* (Upton *et al.*, 2003), which consist of centrally located genus-common genes with most genus-specific genes located terminally. In these families, many genus- and virus-specific genes are involved in interactions with the host, presumably to promote survival in relevant biological niches, and a number have been captured from the host. It is interesting to note that eukaryotes, as represented by yeast, exhibit rapid evolution in the telomeric regions of chromosomes (Kellis *et al.*, 2003).

Gene capture has played an important role throughout adenovirus evolution. Thus, genus-common genes, such as pol, IVa2 (which contains an ATP-binding domain) and protease, possibly resulted from very ancient capture events. Table 3 lists adenovirus genes with homologues in cellular or other viral genomes that have been captured more recently, after the genera diverged. Genes with cellular counterparts are taken to represent imports into adenoviruses, but those with counterparts only in other viruses

Tabl 3. Genus-specific genes captured during adenovirus evolution

Virus	Gene	H m logue	C mments
MASTADENOVIRUS			
Most	E4 ORF1	Cellular dUTPase	Duplicated, active site motifs retained in some
Primate	E3 CR1	Primate cytomegalovirus CR1 protein	Duplicated
ATADENOVIRUS			
All	p32K	Bacterial small acid-soluble protein	Tenuous relationship
All	RH	Cellular F-box protein	Duplicated
DAdV-1	ORF4	Cellular OX-2 glycoprotein	
DAdV-1	ORF7	Cellular RING-finger protein	
AVIADENOVIRUS			
All	ORF1	Cellular dUTPase	Active site motifs retained
All	ORF2	Parvovirus Rep	Duplicated
All	ORF11	Cellular leukocyte adhesion molecule	Duplicated
All	ORF19	Putative avian herpesvirus lipase	
FAdV-1	ORF26	Cellular C-type lectin	
FAdV-10	ORF4	Avian herpesvirus US22 protein	
SIADENOVIRUS			
All	sialidase	Cellular sialidase	

could have been transferred from virus to virus in either direction. Whether discernable as captured or not, genus-specific genes range from two (E1B 55K and E4 34K) whose development pre-dated divergence of mastadenoviruses and atadenoviruses, through genes that characterize all viruses in a genus (e.g. E1A in mastadenoviruses and p32K in atadenoviruses) to genes that are specific to a subset of viruses, or even a single virus, within a genus (e.g. E3 12·5K in mastadenoviruses). This complexity in definition is not unexpected, since genes could be gained or lost at any stage.

The processes of gene duplication, divergence and functional partition have clearly been at work among genus-specific genes. This is an evolutionary mechanism used widely by larger DNA viruses for generating new genes. Several examples in adenoviruses are listed in Table 3, and include CR1 genes in human mastadenoviruses, the dUTPase gene in mastadenoviruses, 34K in mastadenoviruses and atadenoviruses, and RH genes in aviadenoviruses. The involvement of gene duplication in the leader–exon setting of the E3 and E4 regions is discussed below and highlights the utility of this evolutionary strategy. Relatively recent duplications are also apparent in genus-common genes. Examples include fiber in FAdV-1 and HAdV-F and DBP in TSAdV-1. Indeed, PAdV-3 and FrAdV-1 exhibit very recent perfect duplications of sequences near the right end of the genome, which, although possibly a result of passage in cell culture and therefore not of functional or evolutionary significance, nonetheless attest to the ease with which duplications occur. Given the leader–exon structure of late genes, it is tempting to speculate that some of these evolved by ancient duplication events but that any evidence in the form of amino acid sequence similarity has been obliterated over time.

Evolution of E4 in the mastadenoviruses appears particularly complex. In this region, precise genetic relationships are readily catalogued between human adenovirus species, but in other mastadenoviruses the relationships are difficult to assess, although the basic structure and expression pattern appear to be conserved. As a result, genetic relationships between non-primate mastadenovirus E4 genes are not specified in Table 2. In five of the human adenovirus species, E4 contains six leftward-oriented genes (ORF1, ORF2, ORF3, ORF4, 34K and ORF6/7), transcripts from which are regulated by a promoter near the right end of the genome. ORF6/7 mRNA results from further splicing between the 5′ end of 34K and the region immediately downstream of 34K, and thus the ORF6/7 and 34K proteins share N-terminal sequences. HAdV-F is unusual in lacking a counterpart to ORF1, presumably as a result of deletion. ORF1 appears to be derived from a captured dUTPase gene, but its descendents in human adenoviruses have not retained the active site residues and presumably carry out other functions (Weiss *et al.*, 1997). Evolution of E4 in mastadenoviruses has evidently involved duplication or deletion events resulting in variable numbers of genes. Occasionally, alternative routes of functional partition for duplicated 34K or dUTPase-derived genes have led to a situation where sequence similarity between two viruses is greatest between ORFs that do not correspond in location. In atadenoviruses, knowledge of E4 is less extensive, but splicing in leader–exon fashion and gene duplication are evident. Two tandem 34K genes (34K-1 and 34K-2) are present, and the RH family of duplicated genes (termed ORF8 and ORF9 in DAdV-1) encoding F-box proteins has developed upstream (Both, 2002b). Two related genes (ORF5 and ORF6) resulting from duplication are present at the right end of the DAdV-1 genome.

These observations lead to a speculative model for the evolution of E4 in mastadenoviruses and atadenoviruses in which 34K was present in the common ancestor and other genus-specific genes were inserted nearer to the right genome terminus. These included a dUTPase gene in the mastadenovirus lineage and an F-box gene in the atadenovirus lineage. The 34K, dUTPase and F-box genes then proliferated at various stages by duplication, divergence and functional partition, and sometimes by deletion, to give rise, in concert with development of unrelated additional genes, to the E4 formats observed. The dUTPase-derived proteins in modern mastadenoviruses exhibit a range of degrees of relatedness. Most are probably not active dUTPases, but a functional enzyme appears to have been retained in the lineages leading to TSAdV-1 and PAdV-5/BAdV-2. Other clues of gene duplication in E4 also remain, such as in two 34K-derived genes in PAdV-5 and in three genes in TSAdV-1 that are related to ORF2 of primate mastadenoviruses. However, we suspect that divergence of duplicated genes may have proceeded so far in many instances as to obliterate evidence of common origin. It is also interesting to note that duplication of a dUTPase gene has occurred during herpesvirus evolution (McGeoch, 1990), where it is postulated that the initial duplication event resulted in a fused, double-length gene encoding two sets of active site residues. This was followed by loss of one active site to yield a still-active enzyme, and in one lineage by loss of the remaining active site to end in a protein that is presumably no longer a dUTPase (McGeoch & Davison, 1999).

In aviadenoviruses, E4 is substantially larger than in other genera. Understanding of the gene content is incomplete, but useful information on transcription of the FAdV-1 and FAdV-9 genomes (Ojkic *et al.*, 2002; Payet *et al.*, 1998) has aided our extension of previous analyses. It is clear that splicing is common and that gene capture, duplication and functional partition have occurred. A family of putative glycoprotein genes is located near the right genome end in FAdV-1 and FAdV-9. It consists of three rightward-oriented genes (ORF9, ORF10 and ORF11) in the former and one rightward- (ORF11) and one leftward-oriented gene (ORF23) in the latter. The origin of this gene family is revealed by the clear similarity between the ORF11 protein from each virus and cellular leukocyte adhesion molecules. In the absence of obvious cellular forerunners, two observations currently imply that genetic exchange may also have occurred between avian adenoviruses and herpesviruses. Firstly, the ORF19 proteins in FAdV-1 and FAdV-9 are most closely related to membrane proteins (putative lipases) encoded by members of the genus *Mardivirus*, an avian lineage of the *Herpesviridae* (subfamily *Alphaherpesvirinae*) that includes Marek's disease virus (Ojkic & Nagy, 2000). Secondly, partial sequence data indicate that the gene (ORF4) at the right end of the FAdV-10 genome (AF160185) lacks counterparts in FAdV-1 and FAdV-9, but has relatives in several other viruses (Davison *et al.*, 2003). ORF4 is predicted to be spliced, and has closest relatives again in mardiviruses. Distant relatives of this gene are detectable in two other lineages of the *Herpesviridae* (subfamily *Betaherpesvirinae* and an amphibian herpesvirus) and in one lineage of the *Poxviridae* (fowlpox virus).

Gene capture and duplication also feature in E1. An atadenovirus protein (p32K) bears a tenuous relationship to small acid-soluble proteins of bacteria (Élő *et al.*, 2003), and the siadenoviruses are so named because they encode a putative sialidase (Davison & Harrach, 2002). The aviadenoviruses have a dUTPase (ORF1), which retains active site residues, plus a parvovirus Rep protein (ORF2) (Chiocca *et al.*, 1996). Although Chiocca *et al.* (1996) speculated that the dUTPase genes of mastadenoviruses and aviadenoviruses may have evolved via transfer of an ancestral gene from one genome terminus to the other, it is equally likely that they were acquired by separate capture events. Duplicates of aviadenovirus ORF2 appear to be present on the opposing strand (ORF12 and ORF13), and are expressed by splicing from the short coding exon utilized by pTP.

In contrast to E1 and E4, which are present in all genera, E3 features only in mastadenoviruses and siadenoviruses. This location for genus-specific genes may have been arrived at independently, or may represent an ancient locus for rapid gene evolution that has been lost in two genera. A substantial set of genes, numbering up to nine, has evolved in mastadenovirus E3. The two genes at the ends of this block (12·5K and 14·7K) encode distantly related proteins that probably arose via gene duplication at an early stage in mastadenovirus evolution with subsequent loss in some lineages. The intervening genes encode membrane proteins, certain of which (CR1 genes) in primate adenoviruses have relatives in primate cytomegaloviruses (a lineage of subfamily *Betaherpesvirinae* in the *Herpesviridae*). CR1 genes share a common motif but are highly variable in number and sequence in both virus families (Davison *et al.*, 2003). Interestingly, TSAdV-1 lacks CR1 genes in E3, but the cognate domain is present in the putative glycoprotein encoded by a gene (105R-T), which is unique to this virus and situated at the right end of the genome. Both (2002b) detected a weak similarity between the single siadenovirus E3 protein and RH5 of atadenoviruses OAdV-7 and BAdV-4, but the evolutionary significance of this is not clear.

In the genomes of most primate mastadenovirus species (including all human and chimpanzee viruses), the region between the pTP and 52K genes contains one or two rightward-oriented VA RNA genes, which are transcribed by RNA polymerase III (pol III) (Kidd *et al.*, 1995; Ma & Mathews, 1996; Mathews & Shenk, 1991). The encoded partially double-stranded RNAs are approximately 160 nucleotides in size, and are involved in translational control and inhibition of the interferon response (Mathews & Shenk, 1991; Mori *et al.*, 1996). A 90 nucleotide pol III RNA is also produced by a leftward-oriented gene located between ORF16 and ORF9 in FAdV-1 (Larsson *et al.*, 1986). Thus, acquisition of VA RNA genes is likely to have occurred at least twice during adenovirus evolution, once in primate mastadenoviruses (perhaps from a tRNA gene,

followed by gene duplication in some lineages) and once in aviadenoviruses. On the basis of limited sequence similarity to FAdV-1, a candidate VA RNA gene was identified in the DAdV-1 genome, overlapping the 3′ end of ORF4 (Hess *et al.*, 1997). This region is absent from the other two atadenoviruses sequenced, and functional investigations have failed to detect a VA RNA in OAdV-7 (Venktesh *et al.*, 1998). If the identified DAdV-1 VA RNA gene is genuine, its evolution might have been independent or might have involved transfer to or from an aviadenovirus.

Evolution

Similarities between phylogenetic trees for adenovirus protease and the small subunit of host mitochondrial rRNA led to the conclusion that adenoviruses have largely co-speciated with their hosts (Benkő & Harrach, 2003). Mastadenoviruses infect mammalian hosts exclusively, and aviadenoviruses have been found only in birds. The picture becomes blurred for the other two genera. Atadenoviruses have been identified in distantly related hosts, such as various species of poultry, a variety of domestic and wild ruminants, and a marsupial. More recently, atadenoviruses have been shown to be present in snakes and lizards, in support of a reptilian source for this genus (Harrach, 2000). The presence of atadenoviruses in birds and mammals could be explained by a couple of (supposedly independent) host switching events. Although one of the siadenoviruses infects a bird, it is tempting to speculate that this genus corresponds to the original amphibian lineage. The partial genome sequence of a fish adenovirus implies a fifth genus (Fig. 1), thus reinforcing the idea that major vertebrate lineages are characterized by distinct adenovirus genera (Kovács *et al.*, 2003).

The most recent ancestor of all modern adenoviruses is likely have been an adenovirus that existed before the divergence of bony fish from other vertebrates. This virus possessed a substantial prior evolutionary history. Since adenoviruses of invertebrates have not yet been discovered, the characteristics of adenoviruses predating vertebrates are unknown. Nonetheless, tantalizing glimpses may be had into earlier epochs. There are clear similarities in overall architecture of the virion and in the structure of its principal protein component (hexon) between adenoviruses and bacteriophage PRD1 (Belnap & Steven, 2000; Benson *et al.*, 1999, 2002; San Martin & Burnett, 2003). PRD1 belongs to the family *Tectiviridae*, infects Gram-negative hosts and has a linear, double-stranded DNA genome of 15 kbp that is linked to a TP (Bamford, 2002; Bamford & Ackermann, 2000). Adenovirus hexon possesses the 'jellyroll fold' common to capsid proteins of many viruses (Chelvanayagam *et al.*, 1992), but the structural relationship is closest to that of PRD1. Parallels also extend to other features, such as fiber and its PRD1 counterpart (spike), which are present at the virion vertices. These findings provide evidence for divergent evolution of adenoviruses and tectiviruses from an ancestor that pre-dated eukaryotes and exhibited the adenovirus morphology. The ancient origins of this morphology are highlighted further by recent findings on bacteriophage Bam35, which infects a Gram-positive host. PRD1 and Bam35 are morphologically indistinguishable and share the hexon fold, even though they are thought to have diverged over a billion years ago (Ravantti *et al.*, 2003).

Protein-primed DNA replication is employed by adenoviruses and certain bacteriophages (Berencsi *et al.*, 1995; de Jong *et al.*, 2003; Liu *et al.*, 2003; Salas, 1991). The components are a linear double-stranded DNA template with ITRs, pol, TP and at least one DBP. This strategy is used by PRD1 (Bamford *et al.*, 1991) and by bacteriophage ϕ29 (19 kbp), which belongs to the *Podoviridae* and has a different capsid morphology from PRD1 (Pečenková & Pačes, 1999). Moreover, in adenoviruses and these two bacteriophages, TP and pol are early genes arranged in tandem near the left genome end, and the late genes are located more centrally, adjacent to TP and pol and arrayed rightward. The antiquity of protein-primed DNA replication is further underscored by its occurrence in linear plasmids of fungi and plants, which generally encode the required pol and TP (Paillard *et al.*, 1985; Rohe *et al.*, 1991), and by linear bacterial genomes of *Streptomyces* species (Chen, 1996).

In summary, we may speculate that the protein-primed DNA replication strategy originated at a very early stage in evolution, to be followed by acquisition of the adenovirus morphology during a pre-eukaryotic era. Introduction of splicing and additional replicative and structural genes resulted in an adenovirus from which extant members of the family have inherited at least 16 genes. Subsequent lineages developed specific subsets of genes that fit them to particular biological niches.

ACKNOWLEDGEMENTS

This work was supported by the Royal Society, the Hungarian Scientific Research Fund (OTKA T043422), the Hungarian Prime Minister's Office (MEH 4676/1/2003), and the Medical Research Council. We thank Duncan McGeoch for critical reading of the manuscript.

REFERENCES

Akusjärvi, G., Pettersson, U. & Roberts, R. J. (1986). Structure and function of the adenovirus-2 genome. In *Developments in Molecular Virology: Adenovirus DNA*, pp. 53–95. Edited by W. Doerfler. Boston: Martinus Nijhoff.

Anderson, C. W. (1990). The proteinase polypeptide of adenovirus serotype 2 virions. *Virology* **177**, 259–272.

Bailey, A. & Mautner, V. (1994). Phylogenetic relationships among adenovirus serotypes. *Virology* **205**, 438–452.

Baker, C. C. & Ziff, E. B. (1981). Promoters and heterogeneous 5′ termini of the messenger RNAs of adenovirus serotype 2. *J Mol Biol* **149**, 189–221.

Bamford, D. H. (2002). *Tectivirus.* In *The Springer Index of Viruses*, pp. 1132–1135. Edited by C. A. Tidona & G. Darai. Berlin: Springer-Verlag.

Bamford, D. H. & Ack rmann, H.-W. (2000). Family *Tectiviridae.* In *Virus Taxonomy: Classification and Nomenclature of Viruses. Seventh*

Report of the International Committee on Taxonomy of Viruses, pp. 111–116. Edited by M. H. V. van Regenmortel, C. M. Fauquet, D. H. L. Bishop, E. B. Carstens, M. K. Estes, S. M. Lemon, J. Maniloff, M. A. Mayo, D. J. McGeoch, C. R. Pringle & R. B. Wickner. San Diego: Academic Press.

Bamf rd, J. K. H., Hänninen, A.-L., Pakula, T. M., Ojala, P. M., Kalkkinen, N., Frilander, M. & Bamford, D. H. (1991). Genome organization of membrane-containing bacteriophage PRD1. *Virology* **183**, 658–676.

Belnap, D. M. & Steven, A. C. (2000). '*Déjà vu* all over again': the similar structures of bacteriophage PRD1 and adenovirus. *Trends Microbiol* **8**, 91–93.

Benkő, M. & Harrach, B. (1998). A proposal for a new (third) genus within the family *Adenoviridae*. *Arch Virol* **143**, 829–837.

B nkő, M. & Harrach, B. (2003). Molecular evolution of adenoviruses. *Curr Top Microbiol Immunol* **272**, 3–35.

Benkő, M., Harrach, B. & Russell, W. C. (2000). Family *Adenoviridae*. In *Virus Taxonomy. Seventh Report of the International Committee on Taxonomy of Viruses*, pp. 227–238. Edited by M. H. V. van Regenmortel, C. M. Fauquet, D. H. L. Bishop, E. B. Carstens, M. K. Estes, S. M. Lemon, J. Maniloff, M. A. Mayo, D. J. McGeoch, C. R. Pringle & R. B. Wickner. San Diego: Academic Press.

Benkő, M., Élő, P., Ursu, K., Ahne, W., LaPatra, S. E., Thomson, D. & Harrach, B. (2002). First molecular evidence for the existence of distinct fish and snake adenoviruses. *J Virol* **76**, 10056–10059.

Benson, S. D., Bamford, J. K. H., Bamford, D. H. & Burnett, R. M. (1999). Viral evolution revealed by bacteriophage PRD1 and human adenovirus coat protein structures. *Cell* **98**, 825–833.

Benson, S. D., Bamford, J. K. H., Bamford, D. H. & Burnett, R. M. (2002). The X-ray crystal structure of P3, the major coat protein of the lipid-containing bacteriophage PRD1, at 1·65 Å resolution. *Acta Crystallogr D Biol Crystallogr* **58**, 39–59.

Berencsi, G., Bánrévi, A., Takács, M., Lengyel, A. & Nász, I. (1995). Evolutionary aspects, and taxonomic definition of viruses together with mobile extrachromosomal elements of relative autonomy into a special highest rank taxon (a review). *Acta Microbiol Immunol Hung* **42**, 141–153.

Berget, S. M., Moore, C. & Sharp, P. A. (1977). Spliced segments at the 5′ terminus of adenovirus 2 late mRNA. *Proc Natl Acad Sci U S A* **74**, 3171–3175.

Borodovsky, M. & Mclninch, J. (1993). GENMARK: parallel gene recognition for both DNA strands. *Comput Chem* **17**, 123–133.

Both, G. W. (2002a). *Atadenovirus*. In *The Springer Index of Viruses*, pp. 2–8. Edited by C. A. Tidona & G. Darai. Berlin: Springer-Verlag.

B th, G. W. (2002b). Identification of a unique family of F-box proteins in atadenoviruses. *Virology* **304**, 425–433.

Chelvanayagam, G., Heringa, J. & Argos, P. (1992). Anatomy and evolution of proteins displaying the viral capsid jellyroll topology. *J Mol Biol* **228**, 220–242.

Ch n, C. W. (1996). Complications and implications of linear bacterial chromosomes. *Trends Genet* **12**, 192–196.

Chillon, M., Bosch, A., Zabner, J., Law, L., Armentano, D., Welsh, M. J. & Davidson, B. L. (1999). Group D adenoviruses infect primary central nervous system cells more efficiently than those from group C. *J Virol* **73**, 2537–2540.

Chiocca, S., Kurzbauer, R., Schaffner, G., Bak r, A., Mautner, V. & Cotten, M. (1996). The complete DNA sequence and genomic organization of the avian adenovirus CELO. *J Virol* **70**, 2939–2949.

Ch w, L. T., Gelinas, R. E., Broker, T. R. & Roberts, R. J. (1977). An amazing sequence arrangement at the 5′ ends of adenovirus 2 messenger RNA. *Cell* **12**, 1–8.

Ch w, L. T., Brok r, T. R. & L wis, J. B. (1979). Complex splicing patterns of RNAs from the early regions of adenovirus-2. *J Mol Biol* **134**, 265–303.

Chroboczek, J., Bieber, F. & Jacrot, B. (1992). The sequence of the genome of adenovirus type 5 and its comparison with the genome of adenovirus type 2. *Virology* **186**, 280–285.

Davison, A. J. & Harrach, B. (2002). *Siadenovirus*. In *The Springer Index of Viruses*, pp. 29–33. Edited by C. A. Tidona & G. Darai. New York: Springer-Verlag.

Davison, A. J., Telford, E. A. R., Watson, M. S., McBride, K. & Mautner, V. (1993). The DNA sequence of adenovirus type 40. *J Mol Biol* **234**, 1308–1316.

Davison, A. J., Wright, K. M. & Harrach, B. (2000). DNA sequence of frog adenovirus. *J Gen Virol* **81**, 2431–2439.

Davison, A. J., Akter, P., Cunningham, C. & 7 other authors (2003). Homology between the human cytomegalovirus RL11 gene family and human adenovirus E3 genes. *J Gen Virol* **84**, 657–663.

de Jong, R. N., van der Vliet, P. C. & Brenkman, A. B. (2003). Adenovirus DNA replication: protein priming, jumping back and the role of the DNA binding protein DBP. *Curr Top Microbiol Immunol* **272**, 187–211.

Dix, I. & Leppard, K. N. (1992). Open reading frames 1 and 2 of adenovirus region E4 are conserved between human serotypes 2 and 5. *J Gen Virol* **73**, 2975–2976.

Élő, P., Farkas, S. L., Dán, Á. L. & Kovács, G. M. (2003). The p32K structural protein of the atadenovirus might have bacterial relatives. *J Mol Evol* **56**, 175–180.

Farina, S. F., Gao, G.-P., Xiang, Z. Q., Rux, J. J., Burnett, R. M., Alvira, M. R., Marsh, J., Ertl, H. C. J. & Wilson, J. M. (2001). Replication-defective vector based on a chimpanzee adenovirus. *J Virol* **75**, 11603–11613.

Farkas, S. L., Benkő, M., Élő, P., Ursu, K., Dán, Á., Ahne, W. & Harrach, B. (2002). Genomic and phylogenetic analyses of an adenovirus isolated from a corn snake (*Elaphe guttata*) imply a common origin with members of the proposed new genus *Atadenovirus*. *J Gen Virol* **83**, 2403–2410.

Harrach, B. (2000). Reptile adenoviruses in cattle? *Acta Vet Hung* **48**, 485–490.

Harrach, B. & Benkő, M. (1998). Phylogenetic analysis of adenovirus sequences. Proof of the necessity of establishing a third genus in the *Adenoviridae* family. *Methods Mol Med* **21**, 309–339.

Hess, M., Blöcker, H. & Brandt, P. (1997). The complete nucleotide sequence of the egg drop syndrome virus: an intermediate between mastadenoviruses and aviadenoviruses. *Virology* **238**, 145–156.

Honess, R. W., Gompels, U. A., Barrell, B. G., Craxton, M., Cameron, K. R., Staden, R., Chang, Y.-N. & Hayward, G. S. (1989). Deviations from expected frequencies of CpG dinucleotides in herpesvirus DNAs may be diagnostic of differences in the states of their latent genomes. *J Gen Virol* **70**, 837–855.

Kellis, M., Patterson, N., Endrizzi, M., Birren, B. & Lander, E. S. (2003). Sequencing and comparison of yeast species to identify genes and regulatory elements. *Nature* **423**, 241–254.

Khatri, A. & Both, G. W. (1998). Identification of transcripts and promoter regions of ovine adenovirus OAV287. *Virology* **245**, 128–141.

Kidd, A. H., Garwicz, D. & Öberg, M. (1995). Human and simian adenoviruses: phylogenetic inferences from analysis of VA RNA genes. *Virology* **207**, 32–45.

Kovács, G. M., LaPatra, S. E., D'Halluin, J.-C. & Benk", M. (2003). Phylogenetic analysis of the hexon and protease genes of a fish adenovirus isolated from white sturgeon (*Acipenser transmontanus*) supports the proposal for a new adenovirus genus. *Virus Res* (in press).

Larsson, S., Bell tt, A. & Akusjärvi, G. (1986). VA RNAs from avian and human adenoviruses: dramatic differences in length, sequence, and gene location. *J Virol* **58**, 600–609.

Liu, H., Naismith, J. H. & Hay, R. T. (2003). Adenovirus DNA replication. *Curr Top Microbiol Immunol* **272**, 131–164.

Ma, Y. & Mathews, M. B. (1996). Structure, function, and evolution of adenovirus-associated RNA: a phylogenetic approach. *J Virol* **70**, 5083–5099.

Mathews, M. B. & Shenk, T. (1991). Adenovirus-associated RNA and translation control. *J Virol* **65**, 5657–5662.

McGeoch, D. J. (1990). Protein sequence comparisons show that the 'pseudoproteases' encoded by poxviruses and certain retroviruses belong to the deoxyuridine triphosphatase family. *Nucleic Acids Res* **18**, 4105–4110.

McGeoch, D. J. & Davison, A. J. (1999). The molecular evolutionary history of the herpesviruses. In *Origin and Evolution of Viruses*, pp. 441–465. Edited by E. Domingo, R. Webster & J. Holland. London: Academic Press.

Mei, Y.-F., Skog, J., Lindman, K. & Wadell, G. (2003). Comparative analysis of the genome organization of human adenovirus 11, a member of the human adenovirus species B, and the commonly used human adenovirus 5 vector, a member of species C. *J Gen Virol* **84**, 2061–2071.

Meissner, J. D., Hirsch, G. N., LaRue, E. A., Fulcher, R. A. & Spindler, K. R. (1997). Completion of the DNA sequence of mouse adenovirus type 1: sequence of E2B, L1, and L2 (18–51 map units). *Virus Res* **51**, 53–64.

Mori, K., Juttermann, R., Wienhues, U., Kobayashi, K., Yagi, M., Sugimoto, T., Tjia, S. T., Doerfler, W. & Hosokawa, K. (1996). Anti-interferon activity of adenovirus-2-encoded VAI and VAII RNAs in translation in cultured human cells. *Virus Res* **42**, 53–63.

Morrison, M. D., Onions, D. E. & Nicolson, L. (1997). Complete DNA sequence of canine adenovirus type 1. *J Gen Virol* **78**, 873–878.

Nagy, M., Nagy, E. & Tuboly, T. (2001). The complete nucleotide sequence of porcine adenovirus serotype 5. *J Gen Virol* **82**, 525–529.

Ojkic, D. & Nagy, E. (2000). The complete nucleotide sequence of fowl adenovirus type 8. *J Gen Virol* **81**, 1833–1837.

Ojkic, D., Krell, P. J. & Nagy, E. (2002). Unique features of fowl adenovirus 9 gene transcription. *Virology* **302**, 274–285.

Page, R. D. M. (1996). TREEVIEW: an application to display phylogenetic trees on personal computers. *Comput Appl Biosci* **12**, 357–358.

Paillard, M., Sederoff, R. R. & Levings, C. S., III (1985). Nucleotide sequence of the S-1 mitochondrial DNA from the cytoplasm of maize. *EMBO J* **4**, 1125–1128.

Payet, V., Arnauld, C., Picault, J.-P., Jestin, A. & Langlois, P. (1998). Transcriptional organization of the avian adenovirus CELO. *J Virol* **72**, 9278–9285.

Pečenková, T. & Pačes, V. (1999). Molecular phylogeny of ϕ29-like phages and their evolutionary relatedness to other protein-primed replicating phages and other phages hosted by gram-positive bacteria. *J Mol Evol* **48**, 197–208.

Pitc vski, J., Mualem, M., Rei-K r n, Z. & 7 other authors (1998). The complete DNA sequence and genome organization of the avian adenovirus, hemorrhagic enteritis virus. *Virology* **249**, 307–315.

Ravantti, J. J., Gaidelyt , A., Bamford, D. H. & Bamford, J. K. H. (2003). Comparative analysis of bacterial viruses Bam35, infecting a gram-positive host, and PRD1, infecting gram-negative hosts, demonstrates a viral lineage. *Virology* (in press).

Reddy, P. S., Idamakanti, N., Song, J.-Y. & 7 ther auth rs (1998a). Nucleotide sequence and transcription map of porcine adenovirus type 3. *Virology* **251**, 414–426.

Reddy, P. S., Idamakanti, N., Zakhartch uk, A. N., Baxi, M. K., L e, J. B., Pyne, C., Babiuk, L. A. & Tikoo, S. K. (1998b). Nucleotide sequence, genome organization, and transcription map of bovine adenovirus type 3. *J Virol* **72**, 1394–1402.

Roberts, R. J., Akusjärvi, G., Aleström, P., Gelinas, R. E., Gingeras, T. R., Sciaky, D. & Pettersson, U. (1986). A consensus sequence for the adenovirus-2 genome. In *Developments in Molecular Virology: Adenovirus DNA*, pp. 1–51. Edited by W. Doerfler. Boston: Martinus Nijhoff.

Rohe, M., Schrage, K. & Meinhardt, F. (1991). The linear plasmid pMC3-2 from *Morchella conica* is structurally related to adenoviruses. *Curr Genet* **20**, 527–533.

Russell, W. C. (2000). Update on adenovirus and its vectors. *J Gen Virol* **81**, 2573–2604.

Russell, W. C. & Benkő, M. (1999). Animal adenoviruses. In *Encyclopedia of Virology*, pp. 14–21. Edited by A. Granoff & R. G. Webster. New York: Academic Press.

Ruzindana-Umunyana, A., Imbeault, L. & Weber, J. M. (2002). Substrate specificity of adenovirus protease. *Virus Res* **89**, 41–52.

Salas, M. (1991). Protein-priming of DNA replication. *Annu Rev Biochem* **60**, 39–71.

San Martin, C. & Burnett, R. M. (2003). Structural studies on adenoviruses. *Curr Top Microbiol Immunol* **272**, 57–94.

Schöndorf, E., Bahr, U., Handermann, M. & Darai, G. (2003). Characterization of the complete genome of the Tupaia (tree shrew) adenovirus. *J Virol* **77**, 4345–4356.

Shenk, T. (2001). *Adenoviridae*: the viruses and their replication. In *Fields Virology*, 4th edn, vol. 2, pp. 2265–2300. Edited by D. M. Knipe & P. M. Howley. Philadelphia: Lippincott Williams & Wilkins.

Sheppard, M., Werner, W., McCoy, R. J. & Johnson, M. A. (1998). The major late promoter and bipartite leader sequence of fowl adenovirus. *Arch Virol* **143**, 537–548.

Sprengel, J., Schmitz, B., Heuss-Neitzel, D., Zock, C. & Doerfler, W. (1994). Nucleotide sequence of human adenovirus type 12 DNA: comparative functional analysis. *J Virol* **68**, 379–389.

Stone, D., Furthmann, A., Sandig, V. & Lieber, A. (2003). The complete nucleotide sequence, genome organization, and origin of human adenovirus type 11. *Virology* **309**, 152–165.

Tigges, M. A. & Raskas, H. J. (1984). Splice junctions in adenovirus 2 early region 4 mRNAs: multiple splice sites produce 18 to 24 RNAs. *J Virol* **50**, 106–117.

Upton, C., Slack, S., Hunter, A. L., Ehlers, A. & Roper, R. L. (2003). Poxvirus orthologous clusters: toward defining the minimum essential poxvirus genome. *J Virol* **77**, 7590–7600.

Venktesh, A., Watt, F., Xu, Z. Z. & Both, G. W. (1998). Ovine adenovirus (OAV287) lacks a virus-associated RNA gene. *J Gen Virol* **79**, 509–516.

Virtanen, A., Gilardi, P., Näslund, A., LeMoullec, J. M., Pettersson, U. & Perricaudet, M. (1984). mRNAs from human adenovirus 2 early region 4. *J Virol* **51**, 822–831.

Vrati, S., Bro kes, D. E., Strike, P., Khatri, A., B yle, D. B. & Both, G. W. (1996). Unique genome arrangement of an ovine adenovirus: identification of new proteins and proteinase cleavage sites. *Virology* **220**, 186–199.

Wadell, G. (1984). Molecular epidemiology of human adenoviruses. *Curr Top Microbiol Immunol* **110**, 191–220.

Weber, J. M. (1995). Adenovirus endopeptidase and its role in virus infection. *Curr Top Microbiol Immunol* **199**, 227–235.

W bster, A., Russ ll, S., Talb t, P., Russell, W. C. & Kemp, G. D. (1989). Characterization of the adenovirus proteinase: substrate specificity. *J Gen Virol* **70**, 3225–3234.

Weiss, R. S., Lee, S. S., Prasad, B. V. V. & Javier, R. T. (1997). Human adenovirus early region 4 open reading frame 1 genes encode growth-transforming proteins that may be distantly related to dUTP pyrophosphatase enzymes. *J Virol* **71**, 1857–1870.

W ld, W. S. M. & Gooding, L. R. (1991). Region E3 of adenovirus: a cassette of genes involved in host immunosurveillance and virus–cell interactions. *Virology* **184**, 1–8.

Zhang, W. & Imperiale, M. J. (2003). Requirement of the adenovirus IVa2 protein for virus assembly. *J Virol* **77**, 3586–3594.